

Utilization and Extraction of Relation Information Considering Features of Languages and Sentences

著者	肥合 智史
発行年	2021-03-25
その他のタイトル	言語や文の特徴を考慮した関係情報の活用と抽出
学位授与番号	17104甲情工第353号
URL	http://hdl.handle.net/10228/00008297

氏 名	肥 合 智 史
学位の種類	博 士（情報工学）
学位記番号	情工博甲第353号
学位授与の日付	令和3年 3月25日
学位授与の条件	学位規則第4条第1項該当
学位論文題目	Utilization and Extraction of Relation Information Considering Features of Languages and Sentences (言語や文の特徴を考慮した関係情報の活用と抽出)
論文審査委員	主 査 教 授 嶋 田 和 孝 " 岡 部 孝 弘 " 乃 万 司 " 久 代 紀 之 " 齊 藤 剛 史

学 位 論 文 内 容 の 要 旨

言語は文脈を持った語の並びであり、それぞれの語の間には様々な関係がある。この関係を特定することや、利用することは自然言語処理において重要な課題である。たとえば、基盤的な技術として、日本語の場合、文字列の関係を踏まえ、単語に分割する形態素解析やその形態素解析の結果を利用して文の構造を特定する係り受け解析などがある。また単語間の意味的な関係を捉えることも重要な課題である。すなわち、自然言語処理において、語の関係情報の抽出や利用は、言語理解のための重要な研究テーマである。

本論文では、自然言語処理における関係情報の抽出と利用について焦点を当てている。関係情報を用いた言語理解の研究において、言語現象の特徴を考慮した関係情報が言語理解モデルに有用かどうかを皮肉抽出というタスクで検証している。また、文中に生じるタスク特有の特徴量が関係情報の抽出の精度向上に有用かどうかを医化学論文中の関係情報抽出というタスクを通して検証している。

皮肉抽出において、これまでの研究では表層レベルの特徴量を SVM や深層学習などの機械学習技術によってモデル化することがほとんどであった。一方で、皮肉という言語現象には、それ特有の特徴などが存在する。その特徴として、皮肉では立場を表す表現対が存在しやすいという仮定をおき、それを関係ベクトルとして定義している。本論文では、この関係ベクトルを考慮した深層学習モデルの有効性を検証している。関係ベクトルは言語コーパスから自動的に獲得している。また、関係ベクトルをどのように組み込むことが効率的かについて実験的に検証をしている。深層学習に基づく高精度な先行研究と比較し、提案手法が優位であることが実験的に検証された。また、関係ベクトル

ルをどのように組み込むことが重要であるかについての知見も得た。

医化学論文の関係情報抽出タスクでは、化合物などは特徴的な接頭辞・接尾辞を持つ傾向があるという文中に生じるタスク特有の現象に着目し、文字レベルと単語レベルの分散表現ベクトルを組み合わせる手法を提案している。近年の自然言語処理では BERT と呼ばれる大規模なニューラルネットワークモデルが幅広いタスクで大きな成果を上げており、医化学分野の関係情報抽出でも同様である。一方で BERT を基にした手法ではハイエンド GPU といった大規模な計算資源が必要である。本論文で提案されている手法は、文字・単語レベルの統合モデルであり、BERT モデルに比べて軽量である。また、統合モデルであるため、状況に応じて任意のモデルを自由に組み合わせることができるというメリットも存在する。実験では、手法の有効性検証のために、複数のベンチマークデータセットにより、提案手法と BERT や各データセットでの最高精度の手法とを比較している。結果として、提案手法は、小規模でシンプルなタスクにおいては BERT や既存の最高精度の手法よりも高精度であることが確認された。また、実行速度についても BERT と比較して 3 倍程度高速であることも確認された。

学 位 論 文 審 査 の 結 果 の 要 旨

本論文に関して、論文調査及び最終試験を行った。論文審査委員やその他の多数の聴講者から、自然言語処理における本研究の位置づけ、提案手法の新規性や有効性の詳細、今後の展開など、種々の質問がなされたが、いずれの質問に対しても論文提出者からの満足な回答が得られた。

以上により、論文調査及び最終試験の結果に基づき、審査委員会において慎重に審査した結果、本論文が、博士（情報工学）の学位に十分値するものであると判断した。